**Supplementary Materials**

Information Retrieval for Biomedical Datasets:

The 2016 bioCADDIE Dataset Retrieval Challenge

Kirk Roberts, PhD1. Anupama E. Gururaj, PhD1. Xiaoling Chen, PhD1.

Saeid Pournejati, MS1. William R. Hersh, MD2. Dina Demner-Fushman, MD, PhD3. Lucila Ohno-Machado, MD, PhD4. Trevor Cohen, MBChB, PhD1. PhD. Hua Xu, PhD1.

1. The University of Texas Health science Center at Houston, Houston, TX.

2. Oregon Health and Science University, Portland, OR.

3. U.S. National Library of Medicine, Bethesda, MD.

4. University of California San Diego, La Jolla, CA.

Corresponding Authors:

Kirk Roberts, [kirk.roberts@uth.tmc.edu](mailto:kirk.roberts@uth.tmc.edu)

Trevor Cohen, [trevor.cohen@uth.tmc.edu](mailto:trevor.cohen@uth.tmc.edu)

Hua Xu, [hua.xu@uth.tmc.edu](mailto:hua.xu@uth.tmc.edu)

Supplementary Table 1: Formatted JSON representation of dataset from first row of Table 1.

|  |
| --- |
| Title: Effect of Meal Replacement Beverage(Glucerna) on Body Composition,Lipid Metabolism and Blood Pressure Metadata:  {  "clinical\_study": {  "oversight\_info": {  "authority": "China: Food and Drug Administration",  "has\_dmc": "Yes"  }  },  "dataItem": {  "dataTypes": ["Study", "Dataset", "StudyGroup", "DataSet", "clinical\_study", "internal", "Treatment", "Grant", "Disease"]  },  "StudyGroup": {  "type": ["Experimental"],  "description": ["Glucerna 52g instead of night meal 5weeks"],  "name": ["Glucerna,Meal Replacement"]  },  "Grant": {  "funder": "Shanghai 10th People's Hospital"  },  "Study": {  "status": "Completed",  "recruits": {  "gender": "Both",  "minimum\_age": "14 Years",  "maximum\_age": "65 Years",  "criteria": "Inclusion Criteria: 1. age\uff1a14~60years 2. BMI \u2265 28kg/m2  3. without liver, kidney, gastrointestinal and other major  organic serious diseases,lactating women, pregnancy"  },  "phase": "N/A",  "identifier": "Glucerna",  "homepage": "https://clinicaltrials.gov/show/NCT02118389",  "studyType": "Interventional"  },  "Disease": {  "name": "Obesity"  },  "DataSet": {  "identifier": "NCT02118389"  },  "internal": {  "link\_text": "Link to the current ClinicalTrials.gov record.",  "rank": "SCR:002309"  },  "Treatment": {  "description": "Glucerna 52g meal replacement",  "agent": "Glucerna",  "title": "Dietary Supplement"  },  "Dataset": {  "briefTitle": "Effect of Meal Replacement Beverage(Glucerna) on Body Composition,Lipid Metabolism and Blood Pressure",  "is\_fda\_regulated": "Yes",  "verificationDate": "October 2013",  "creator": "Shanghai 10th People's Hospital",  "title": "The Effect of Meal Replacement Beverage(Glucerna) on Body Composition,Lipid Metabolism and Blood Pressure in Patients  With Obesity",  "releaseDate": "April 17, 2014",  "has\_expanded\_access": "No",  "depositionDate": "April 1, 2014",  "description": "Effect of Meal Replacement Beverage(Glucerna) on Body Composition,lipid metabolism and blood pressure"  }  } |

Supplementary Table 2: Formatted JSON representation of dataset from second row of Table 1.

|  |
| --- |
| Title: Locality Info and Genetic and Phenotypic Data for Western Scrub-Jay Museum Specimens Metadata:  {  "datastandard": {  "homepage": "dublincore.org",  "name": "Dublin Core",  "license": http://dublincore.org/about/copyright/  },  "dataItem": {  "dataTypes": ["dataset", "internal", "identifiers", "dataRepository", "organization", "datastandard"]  },  "dataRepository": {  "homepage": "http://www.datadryad.org",  "ID": "SCR:005910",  "name": "Dryad Data Repository"  },  "identifiers": {  "ID": ["doi:10.5061/dryad.57f48/1", "http://hdl.handle.net/10255/dryad.64554"]  },  "dataset": {  "dateAvailable": "20140619T153456+0000",  "description": "This Excel spreadsheet lists all 689 Western Scrub-Jay museum specimens used in the study. Museum abbreviations:   Field Museum of Natural History (FMNH); Moore Laboratory of Zoology, Occidental College (MLZ); Museum of Vertebrate   Zoology, Berkeley (MVZ). Haplotypes of the cytochrome B gene are coastal (C) or interior (I). Structure results refer to the   raw results making up the different panels of Figure 3. The raw microsatellite data are listed in actual allele sizes after   allele calling. For the phenotypic data: collar, vent, and eyestripe are qualitatively-scored categorical data ranging from 1   to 6. Collar is the amount of blue coming down from the shoulders under the necklace with 1 = heavy collar (coastal) and   6 = little collar (interior). Vent is the amount of blue tinging to the feathers under the tail coverts, with 1 = white feathers   (coastal) and 6 = heavily blue-tinged feathers (interior). Eyestripe is the boldness of the white stripe above the eye, with 1   = bold stripe (coastal) and 6 = small stripe (interior). Other phenotypic data are in the form of continuous variables   measured in millimeters. Blank cells mean there is no data taken for that individual at that variable.",  "license": "http://creativecommons.org/publicdomain/zero/1.0/",  "title": "Locality Info and Genetic and Phenotypic Data for Western Scrub-Jay Museum Specimens",  "ID": "oai:datadryad.org:10255/dryad.64554",  "downloadURL": "http://www.datadryad.org",  "dateIssued": "20140619",  "keywords": ["birds", "speciation", "phylogeography", "post-zygotic reproductive barriers", "gene flow"],  "dateLastUpdate": "20140619",  "dateAccession": "20140619T153456+0000",  "relatedDataset": "hdl:8"  },  "internal": {  "recNum": "99",  "setID": "hdl\_10255\_2",  "type": "Dataset",  "setName": "Main"  },  "organization": {  "abbreviation": "Dryad",  "name": "Dryad Digital Repository",  "ID": "SCR:005910"  }  } |

Supplementary Table 3: Formatted JSON representation of dataset from third row of Table 1.

|  |
| --- |
| Title: NATIVE CARDOSIN A FROM CYNARA CARDUNCULUS L. Metadata:  {  "dataResource": {  "keywords": [],  "altNames": [],  "acronyms": []  },  "citation": {  "DOI": "doi:10.1074/jbc.274.39.27694",  "author": {  "name": ["Frazao, C.", "Bento, I.", "Costa, J.", "Soares, C.M.", "Verissimo, P.", "Faro, C.", "Pires, E.", "Cooper, J.", "Carrondo, M.A."]  },  "journal": "J.Biol.Chem.",  "title": "Crystal structure of cardosin A, a glycosylated and Arg-Gly-Asp-containing aspartic proteinase from the flowers of Cynara   cardunculus L.",  "journalISSN": "0021-9258",  "firstPage": "27694",  "lastPage": "27701",  "year": "1999",  "PMID": "pmid:10488111"  },  "materialEntity": [  {  "formula": "C2 H5 N O2",  "role": "chemical component",  "name": "GLYCINE",  "weight": "75.067",  "type": "peptide linking"  }, {  "formula": "C3 H7 N O3",  "role": "chemical component",  "name": "SERINE",  "weight": "105.093",  "type": "L-peptide linking"  }, {  < *22 more* >  }, {  "formula": "H2 O",  "role": "chemical component",  "name": "WATER",  "weight": "18.015",  "type": "non-polymer"  }  ],  "dataItem": {  "keywords": ["HYDROLASE", "HYDROLASE", "ASPARTIC PROTEINASE"],  "dataTypes": ["citation", "materialEntity", "dataItem", "identifiers"],  "title": "NATIVE CARDOSIN A FROM CYNARA CARDUNCULUS L.",  "description": "PROTEIN (CARDOSIN A) (3.4.23.-)",  "ID": "1B5F"  },  "identifiers": [  {  "ID": "pdb:1B5F"  }, {  "ID": "rcsb:RCSB008008"  }  ]  } |